DOCUMENT RESUME

ED 151 384 TM 006 802

AUTHOR Filby, Nikola N.
TITLE Theoretical and Empirical Consideration of Design Issues Related to Experimental Teacher Units. Beginning Teacher Evaluation Study. Technical Note Series. Technical Note V-3 For Phase III-A Continuation.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY California State Commission for Teacher Preparation and Licensing, Sacramento.; National Inst. of Education (DHEW), Washington, D.C.
REPORT NO BTES-TN-5-3
PUB DATE Jun 77
NOTE 18p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Achievement Tests; Curriculum Planning; Educational Objectives; Effective Teaching; *Elementary School Curriculum; Elementary School Mathematics; Elementary School Teachers; Grade 2; Primary Education; *Program Length; Reading; *Research Design; Sampling; *Teacher Behavior; Teacher Evaluation; Time Factors (Learning); *Units of Study

ABSTRACT

Alternative theoretical approaches influencing the research design of the Beginning Teacher Evaluation Study are discussed. One line of research used an Experimental Teaching Unit (ETU) to create a semi-standardized situation for comparison of teachers. Though this approach introduces a uniformity which appears to increase comparability, it is at the cost of a somewhat unrepresentative sample of teacher behavior. Another line of research focused on naturally-occurring behavior, particularly instructional time. The design chosen for the Beginning Teacher Evaluation Study involved an "objectives-only" experimental teaching unit incorporated into the instructional time framework. Specific objectives in reading and mathematics were taught, but no materials were provided--only objectives. The effectiveness of 6-, 10-, and 40-day instructional programs were compared. The 40-day period was found to be most favorable in terms of consistent gains and variable amounts of actual instructional time. This method was also recommended because it provided a substantial sample of instruction, yet allowed for more than one sample during the school year. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151384

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BTES

BEGINNING TEACHER EVALUATION STUDY

Technical Note Series

Technical Note V-3
For Phase III-A Continuation

THEORETICAL AND EMPIRICAL CONSIDERATION OF DESIGN ISSUES RELATED TO EXPERIMENTAL TEACHER UNITS

Nikola N. Filby

June, 1977

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 FOLSOM STREET • SAN FRANCISCO CALIFORNIA 94103 • (415) 565-3000

TM 006802

Technical Note V-3
For Phase III-A Continuation

THEORETICAL AND EMPIRICAL CONSIDERATION OF DESIGN ISSUES RELATED TO EXPERIMENTAL TEACHING UNITS

Nikola N. Filby

Submitted to

California Commission for Teacher Preparation and Licensing
1020 "O" Street
Sacramento, California 95814

June 1977

Far West Laboratory for Educational
Research and Development
1855 Folsom Street
San Francisco, California 94103

PREFACE

This is one of a series of reports by the Far West Laboratory for Educational Research and Development, concerned with the study of teacher effectiveness. Under contract to the California Commission for Teacher Preparation and Licensing, with funding provided by the National Institute of Education, the Laboratory is studying teachers in second and fifth grade classes in order to identify teacher behaviors and classroom qualities that are related to reading and mathematics achievement. This multi-year research effort is titled the Beginning Teacher Evaluation Study (BTES). The study addresses a major policy issue of the California Commission which is the agency charged with certifying the appropriateness of teacher training programs throughout the State. Simply put, the Commission wants information about what teacher behaviors are related to student outcomes. This information will then be used jointly by the Commission and the State institutions that it certifies in order to better insure that beginning teachers receive training in areas that have been empirically demonstrated to affect student learning.

This Technical Note discusses some theoretical and empirical issues that influenced the selection of a research design for the major field study carried out in Phase III-B of BTES. One line of research in the hypothesis-generation phase of the study used an Experimental Teaching Unit (ETU) to create a semi-standardized situation for comparison of teachers. Another line of research focused on naturally-occurring behavior, particularly instructional time. In the instructional time framework, a different research design was used, involving an "objectives-only" ETU.

As plans were made for the major field study, design issues were reconsidered. Technical Note V-2 (Filby, 1976) stated the need to reconsider the reasons for an experimental teaching unit and to evaluate the alternative design approaches that had been used thus far. This Technical Note reviews and discusses issues of research design, presents data from alternative approaches, and draws conclusions for the Phase III-B design.

Nikola N. Filby
Program Associate
Beginning Teacher Evaluation Study
June, 1977

Theoretical Considerations

Research studies on instructional processes involve variations on one standard research design. Achievement is measured at two points in time. Gain in achievement is related to various aspects of instruction in the intertest period. Studies differ in the length of the intertest period, the content and specificity of the achievement measures, and the variables used to describe the intertest period. Most commonly, a standardized achievement test is given to measure year-long growth, and psychological or sociological process measures are used to describe the year's instruction.

One problem with this common design is that it ignores content coverage in the classroom. While all classes teach reading and mathematics, classes differ in the specific skills taught and in the amount of time spent on different skills. If one class does not cover multiplication and another class spends two months on this skill, then different achievement measures are needed to assess growth.

Another problem arises from the full-year time span. To produce growth over a full year, a myriad of factors are involved and extensive observation is needed to represent the period fully. All of these factors compete with each other to predict one growth score. Besides the complexity of untangling causation, this is an inefficient design. If achievement is measured over a shorter time period, then the initial study and one or more replications or follow-up studies can all be carried out within one academic year.

Known-sample ETU's. In Phase III-A of the Beginning Teacher Evaluation Study, Experimental Teaching Units (ETU's) were used as part of the research plan (Berliner, Filby, Marliave, Moore, Tikunoff, 1976). Each unit was

a two-week curriculum package. The unit had a set of specific instructional objectives, and all teachers taught toward objectives in that set. The overall amount of instructional time was also fixed. At grade two, each teacher was asked to spend 25 minutes per day on the ETU. Within this general framework, the teacher was encouraged to adapt the unit to the needs of her particular students. Teachers could spend varying amounts of time on different objectives within the unit; they could teach objectives in any order they wanted; they could use sample materials from the unit or any other materials they had available.

Considering the issues presented above, the known sample ETU is a distinct alternative to the more common research design. Content coverage is controlled at an intermediate level of specificity. All grade two classes teach measurement and they all teach it for the same length of time. The period between tests is short -- only two weeks.

This semi-standardized situation involves a delicate balance between control and artificiality. To ask all teachers to teach in the same content area at the same time introduces a uniformity which appears to increase comparability, but at the cost of a somewhat unrepresentative sample of behavior. Teachers are accustomed to proceeding through an orderly curriculum sequence, where they assume that one thing must be taught before another, as planned by the expert curriculum designer. Teacher planning has to do with the management and pacing of this sequence. An ETU disrupts this sequence by requiring a specific content at a specific time. This content may be more or less appropriate for different students. Also, the role of the teacher as planner is substantially altered in the ETU methodology.

Instructional Time Theory. Through a process of literature review and conceptual analysis, the Far West Lab staff conducting BTES has developed a theory centered around academic learning time. In summary, the theory states that student behavior is a necessary mediator between instructional process and measures of achievement outcomes. In particular, engaged learning time provides a measure of learning as it takes place. This measure should predict achievement and should provide a dependent variable for evaluating the role of instructional process variables.

This theory deals with the issue of content coverage in a manner directly opposite from the known-sample ETU's. The ETU establishes controls so that content coverage is the same for all teachers. The instructional time theory capitalizes on naturally-occurring variability in content coverage and builds time into the theory itself. From an instructional time viewpoint, variability in content coverage is desirable because it creates a favorable situation for documenting over a relatively short time the role of learning time in producing achievement. The known sample ETU minimizes variability in content coverage; the instructional time theory benefits from maximum variability.

Notice one practical benefit to the instructional time approach: There is little or no disruption of the teacher's normal behavior. The known-sample ETU was frustrating for some teachers because of the requirements it imposed and because of the amount of concentrated planning that was necessary. With the instructional time approach it is possible to let the teacher continue with her normal sequence of instruction. (At least this is possible as long as there is naturally-occurring variability in learning time.)

The "objectives-only" ETU in the Instructional Time Study. In spite of this difference in general approach to the content coverage issue, some aspects of an ETU are desirable within the instructional time theory framework.

One useful component is the incorporation of specific content areas. It is possible to relate total instructional time in mathematics to achievement on a general mathematics test, such as a standardized achievement test. The relationship will appear strong only if the areas of mathematics instruction match the areas covered on the test. Content coverage is still a critical issue. Content coverage is less of a source of error and the instructional time theory can be tested more clearly when more specific content areas are used. Time in multiplication can be compared to achievement in multiplication. While differences in activities are still not accounted for, at least all the instructional time is relevant to the outcome measure that is used.

Specificity of content can be thought of as a continuum. At one end are global definitions such as "reading" versus "mathematics." At the other end are very precise objectives, such as "knowledge of sums through 10." Probably it is most useful to deal with an intermediate level of specificity. This would allow for relatively continuous measures of achievement, with substantial gains but less severe ceiling effect problems. It makes record keeping a more manageable task. It also allows for substantial instructional time in the content area. The disadvantage is that the match between content of instruction and test content is relatively less precise. For this reason, it might be useful to include a few more specific areas, especially where instructional time is likely to be high.

A second useful feature of the ETU is the use of time periods shorter than a full academic year. It is advantageous to be able to get more (or better) information from a year's study. With shorter time periods, a replication can be built into the design. In addition, for the instructional time framework a shorter period between tests reduces differences

due to the point in time within the period at which a particular content was covered, thus providing better control.

Two factors may produce a lower bound on appropriate intertest periods -- amount of gain and amount of time. In order to show statistically powerful relationships, it is important to have relatively large gains to predict and it is important to have generally large but variable amounts of instructional time. Both of these factors are likely to increase over longer time periods.

The term "objectives-only ETU" has been used to describe a design incorporating ETU features into an instructional time framework. One such design was used in the Instructional Time Study carried out in Phase III-A Extension of the Beginning Teacher Evaluation Study (Fisher, 1976a and b). Several content areas were identified for study. These content areas were generally of intermediate specificity within the domains of reading and mathematics. Teachers were told that these were the areas of study, and they were asked to give some instruction in these areas if it seemed appropriate to them to do so. Two different time intervals were included in the study -- two weeks and eight weeks. Pretests and posttests were given in each content area at the beginning and end of each interval. The goal of the study was a pilot analysis of the relationship between instructional time and student achievement. The design is an ETU in the sense that specific content areas were studied over relatively short-term intervals. It is an "objectives only" situation because no materials were provided -- only objectives to cover. In keeping with the instructional time focus, time was allowed to vary both within and across classes.

Conclusions. It is clear that a thorough study of instructional processes must deal with the variability in content coverage across classes. The known sample ETU deals with this issue by standardizing content coverage for a brief experimental period. Such a unit is probably best suited for investigation of specific instructional variables built into the unit. By the very nature of the controls imposed, the unit limits the representativeness of teacher behavior in comparison to the full range of normal full year instruction. As an alternative to the known sample ETU, instructional time theory provides a model for analyzing naturally-occurring variability in content coverage. This approach allows the teacher to carry on his/her normal sequence of curriculum-based instruction. Since the Phase III-B field study is intended to investigate a wide range of instructional variables which influence achievement in California classrooms, the instructional time theory offers the most promising approach for dealing with content coverage.

Within an instructional time framework, it is useful to retain the features of an "objectives-only" ETU. Student achievement tests should include content areas of intermediate specificity within the general areas of reading and mathematics. Time periods less than a full year should be used, if possible, to increase the amount of information derived from the study.

Data on Amounts of Time and Gain

Both the Instructional Time Study and the known-sample ETU's provide data on a key design issue -- the best length of time for the period between tests. The known sample ETU's used a two-week time period. The Instructional Time Study used both a two-week and an eight-week period. The two-week period

increases efficiency, but only if it is a large enough slice of time to give something substantial to look at. Student achievement gains need to be large enough to detect differences reliably. Amounts of instructional time need to be both large enough and variable enough to provide a useful predictor. Both the known-sample ETU's and the Instructional Time Study provide data on the size of achievement gains and on amounts of instructional time. These data are reported in this section.

The known-sample ETU in grade two reading had two sections. One was on word structure, including compound words, prefixes and suffixes. The other section was on following directions. Teachers were allowed to spend time on one or the other or both of these two sections. Ten instructional days were spent on the unit, with 25 minutes of instruction per day. Teachers kept records during instruction and indicated which objectives were covered each day. Since the total amount of instruction is fixed, these records can be used to estimate the total amount of time spent on each objective and on each general section of the unit. An achievement test was given just before and just after the unit was taught. These test scores show the amount of gain over the ten days of instruction.

The Instructional Time Study was also conducted with grade two students and covered some of the same content areas. A large battery of tests was given in October and again in December. The period between tests was approximately 40 days. Teachers kept daily records of the content of instruction within time blocks during this period. These teacher records yield an estimate of the amount of instructional time spent on specific content areas over the 40-day period. In addition, a few content areas were selected for study over a shorter period within the 40 days. Three areas in reading and three in mathematics were selected and parallel tests

were developed. These tests were given at the beginning and end of a two-week period, with approximately six days of instruction between tests. Direct observation was carried out during this six-day period. The teacher records can also be analyzed for allocated time over the six-day period as a subset of the 40 days.

One specific content area in grade two reading can be traced through all these sources of data. Compound words (identification of parts in words like "sailboat") was included in the word structure section of the known-sample ETU. Compound words was also one of the three content areas studied in the six-day subset of the Instructional Time Study, as well as the 40-day full study. Data from all three sources are shown in Table 1.

The first row in Table 1 shows compound word data from the known sample ETU, where ten instructional days were included in the period of study. These data are based on 14 classes in the known sample for which complete data are available. For these 14 classes, the average gain, an unweighted average of class means, was 9%. The average amount of instructional time accumulated over the 10-day period was 25 minutes. That is, for an average student in an average class, a total of 25 minutes was allocated to compound words in the course of the ETU. (This amounts to one full day of instruction out of the 10 in the ETU period.) In the lowest class, no student received any instruction in compound words. The highest class average was 54 minutes of instruction.

The next two rows in Table 1 show comparable data from the Instructional Time Study. These data are based on six grade two classes. It can be seen that as the time period of the study is lengthened, the amount of instructional time increases, as does the gain in student achievement. For this one specific area of compound words, gains are relatively high for a relatively

small amount of time. Even at six days of naturally occurring instruction, there is variability in time and measurable gain. But these factors are considerably stronger for the 40-day period.

Table 2 shows data for the more general area of total word structure. This area includes compound words along with identification of root words, meaning and use of prefixes and suffixes. No general test of word structure was given over the six-day interval in the Instructional Time Study. It can be seen in Table 2 that more time was spent on this general area than for the more specific subset. The two-week known-sample data show substantial time and reasonably good gain. Still, even better time and gain were obtained with naturally-occurring instruction in the 40-day period.

Tables 3 and 4 show data from the Instructional Time Study for which there is no comparable data from the known-sample ETU's. (In grade two mathematics, the ETU dealt solely with measurement and measurement was not included in the Instructional Time Study tests.) These data show one potential problem with short time periods -- unreliable test scores. It seems unlikely that negative gain scores with instruction reflect real loss of knowledge. More likely, students react negatively to frequent repeated testing. Students may be more or less likely to guess; they may read items more or less carefully. Substantial real differences in knowledge are needed to outweigh these measurement problems.

Conclusions. Taking all these data together, it appears that six instructional days is too short to be relied upon for consistently useable data. The ten days of the ETU, with its constraint on content, holds an intermediate position. The 40-day period is most favorable in terms of consistent gains and large, variable amounts of instructional time. A

period like this would be useful in studies like that conducted in Phase III-8. It provides a substantial sample of instruction and allows for more than one sample from the full academic year.

TABLE 1

Compound Words: Time and Achievement Data

Source of Data	Time Period	Average Percent Gain	Allocated Time (minutes)		
			Average	Lowest	Highest
Known-Sample ETU	10 Days	9%	25	0	54
Instructional Time Study	6 Days	5%	20	0	58
Instructional Time Study	40 Days	23%	54	9	131

TABLE 2

Word Structure: Time and Achievement Data

Source of Data	Time Period	Average Percent Gain	Allocated Time (minutes)		
			Average	Lowest	Highest
Known-Sample ETU	10 Days	10%	115	0	162
Instructional Time Study	6 Days	--	37	3	126
Instructional Time Study	40 Days	15%	200	60	629

TABLE 3

Grade 2 Mathematics: Time and Achievement Data
from the Instructional Time Study

Content	Time Period	Average Percent Gain	Allocated Time (minutes)		
			Average	Lowest	Highest
Place Value	6 Days	9%	49	0	157
	40 Days	22%	122	15	272
Subtraction Without Regrouping	6 Days	1%	51	15	92
	40 Days	15%	340	109	546

TABLE 4

Grade 2 Reading: Time and Achievement Data
from the Instructional Time Study

Content	Time Period	Average Percent Gain	Allocated Time (minutes)		
			Average	Lowest	Highest
Long Vowels	6 Days	-3%	30	8	77
	40 Days	11%	298	170	497
Total Decoding	6 Days	-4%	184	73	262
	40 Days	10%	1082	524	1491

REFERENCES

Berliner, D., Filby, N., Marliave, R., Moore J., and Tikunoff, W. Experimental teaching units and the identification of a special sample of classrooms for conducting research on teaching. (BTES Technical Report 76-12-1). San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Filby, N. Plans for further development of Experimental Teaching Units. (BTES Technical Note V-2). San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Fisher, C. A study of instructional time in grade 2 mathematics. (BTES Technical Report II-3). San Francisco: Far West Laboratory for Educational Research and Development, 1976a.

Fisher, C. A study of instructional time in grade 2 reading. (BTES Technical Report II-4). San Francisco: Far West Laboratory for Educational Research and Development, 1976b.